PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-029776
(43)Date of publication of application : 31.01.2003

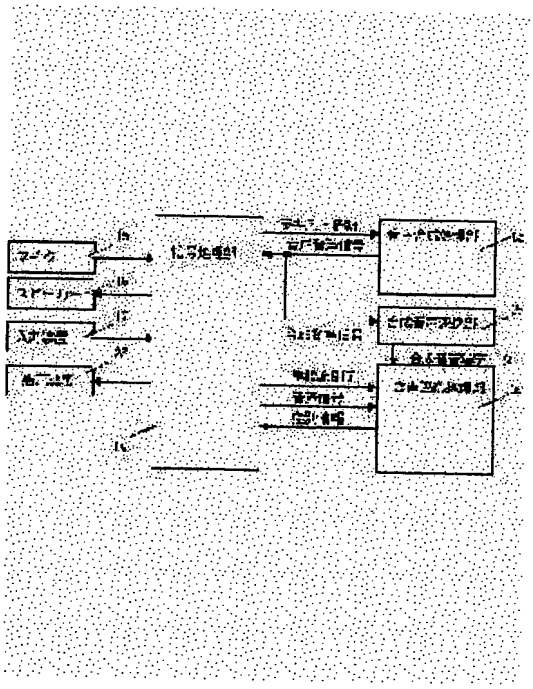
(51)Int.Cl. G10L 15/00
G10L 13/00
G10L 15/06

(21)Application number : 2001-211921
(22)Date of filing : 12.07.2001

(71)Applicant : MATSUSHITA ELECTRIC IND CO LTD
(72)Inventor : NAKAMURA KENJI
OGATA YOSHIYUKI
TATEYAMA MASAKAZU
NISHIDA HIROTO
KUROKI YOSHIAKI
NISHIOKA YASUYUKI
GOSHIMA TATSUHIRO

(54) VOICE RECOGNITION DEVICE

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a voice recognition device which keeps a recognition performance close to the performance of a specific speaker system and is enhanced in convenience by automatically conducting the training.
SOLUTION: When recognition words are registered, text information of the recognition words is inputted into a voice synthesis processing section 15, synthesized voice signals that are to be outputted are converted into synthesized sound acoustic data by a synthesized voice sound converting section 19. The data are registered in a word acoustic data storage section 21 in place of the uttering of a speaker in a conventional training. The coincident process for acoustic data in a word recognition section 26 in the voice recognition process, that is conducted after the registration, is conducted for the synthesized sound acoustic data. Thus, the voice recognition device for a specific speaker is realized at a low cost and the convenience which is similar to the convenience of an unspecified speaker system voice recognition device that does not required training can be provided.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（A）　(11)特許出願公開番号
特開2003－29776
（P2003－29776A）
(43)公開日　平成15年１月31日(2003.1.31)

(51)Int.Cl.'	識別記号	FI	テーマコード'(参考)
G10L 15/00		G10L 3/00	551A 5D015
13/00			521C 5D045
15/06			521B
			R

審査請求　未請求　請求項の数７　ＯＬ　（全 11 頁）

(21)出願番号　特願2001－211921(P2001－211921)

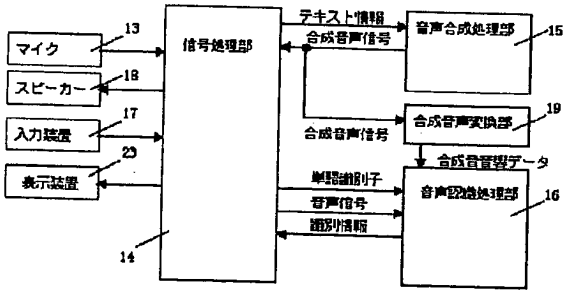
(22)出願日　平成13年７月12日(2001.7.12)

(71)出願人　000005821
松下電器産業株式会社
大阪府門真市大字門真1006番地
(72)発明者　中村　賢二
大阪府門真市大字門真1006番地　松下電器産業株式会社内
(72)発明者　緒方　芳幸
大阪府門真市大字門真1006番地　松下電器産業株式会社内
(74)代理人　100097445
弁理士　岩橋　文雄　（外２名）

最終頁に続く

(54)【発明の名称】　音声認識装置

(57)【要約】
【課題】　本発明は特定話者方式に近い認識性能を保ち、トレーニングを自動的に行うことによって利便性を高めた音声認識装置を提供することを目的とする。
【解決手段】　認識単語の登録の際には、音声合成処理部１５に認識単語のテキスト情報を入力し、出力である合成音声信号を合成音声変換部１９によって合成音音響データに変換し、この合成音音響データを従来のトレーニングによる話者の発声の代わりに単語音響データ格納部２１に登録する。登録後に行われる音声認識処理における単語識別部２６での音響データの一致処理は、この合成音音響データに対して行われる。これにより低コストで実現できる特定話者の音声認識装置を、トレーニングを必要としない不特定話者方式の音声認識装置と同様の利便性を提供することができる。



【特許請求の範囲】

【請求項１】音声入力装置であるマイクと、音声出力装置であるスピーカーと、キーボードなどの入力装置と、認識結果を表示する表示装置と、前記マイク、前記スピーカー、前記入力装置および前記出力装置が接続され、音声認識装置全体の処理制御を実施する信号処理部と、前記信号処理部よりテキスト情報を入力され合成音声信号を出力する音声合成処理部と、前記信号処理部より入力された音声信号を、内部で保持している複数の音響データと比較し、その一致結果を音声認識の結果として前記信号処理へ出力する音声認識処理部を備えた音声認識装置において、前記音声合成処理部からの合成音声信号を音響データである合成音音響データに変換する合成音声変換部を備えることで、合成音音響データを従来は話者によるトレーニングで生成されていた音響データの代わりに音声認識処理部内に保持し音声認識に用いることにより、話者の負担になるトレーニングを必要としないことを特徴とする音声認識装置。

【請求項２】請求項１の音声認識処理部において、合成音声変換部より出力された合成音音響データかあるいは音響処理部より出力された音響データを選択する音響データ選択部を有することにより、単語音響データ格納部に対して装置の初期時は合成音声音響データを格納しておき、ある認識単語を認識した場合には該当単語の発話音声に相当する音響データで合成音音響データを置き換えることで初回以降は実発音での音響データに対する認識を可能とし、合成音声と実発音が異なる場合にも音声認識率の低下を防ぐことを特徴とする音声認識装置。

【請求項３】請求項２の単語音響データ格納部において、前記合成音声変換部より出力された合成音音響データと、前記音響処理部より出力された話者の発話音声の音響データを両方を保持し、両者のうちのいずれかに一致したときに該当する単語の識別情報を出力することで、音声認識率の向上を図ることを特徴とする音声認識装置。

【請求項４】請求項１の単語音響データ格納部において、一つの単語に対して複数の発音の仕方を前記合成音声変換部より入力し格納する構成を持ち、話者がいずれの発音を行った場合でも、該当する単語を正しく認識できることを特徴とする音声認識装置。

【請求項５】請求項４の単語音響データ格納部において、話者が該当する単語を発声し、一致した音響データを残して他の音響データを削除することによって、次回から不要な識別処理を省略し、認識処理をより高速に行うことを特徴とする音声認識装置。

【請求項６】請求項５の単語音響データ格納部において、ある単語に対する複数の合成音音響データのそれぞれに話者の発生が一致した頻度を保持する機構を追加し、該当単語の認識が行われた際に、スレッシュレベル以下の一致頻度の音響データのみを削除することを特徴とする音声認識装置。

【請求項７】請求項１の発明において、個々の識別対象単語に対する合成音響データを音声認識処理部に保持する際に、話者に対して合成音を前記スピーカーより再生し、話者の意図する合成音であるかどうかを確認し、意図しない場合に限り話者によるトレーニング手続きを行うことを特徴とする音声認識装置。

【発明の詳細な説明】

【発明の属する技術分野】本発明は音声認識装置に関するものであり、特定話者を対象とする音声認識技術において、話者によるトレーニング手続きなしで音声認識を行うことを特徴とする音声認識装置に関する。

【従来の技術】近年、電話機やＦＡＸ、カーナビゲーションシステムなどの情報処理装置において、いわゆる音声認識技術を応用して音声入力による本体操作が可能な装置が製品化されるようになってきた。音声認識技術の方式には、話者を限定しない不特定話者方式（speaker independent）と、話者を限定する特定話者方式（speaker dependent）の二つに大別される。不特定話者方式は、音声に含まれる言語的な特徴を抽出し、ニューラルネットワークに代表されるパターン認識技術を応用して話者の発話内容を推定するものである。ところが、話者の発話音声には各個人特有の声質があり、不特定の話者に対して安定した認識率を確保するためには、複雑な処理を必要とする。結果として製品のコストアップにつながる。一方、特定話者方式は対応できる話者を限定することにより安価なシステムで良好な音声認識率を得るものである。この方式では、装置の初回使用時に話者自身の声質を登録（トレーニング）することが必要であり、その分の手間を必要とする。音声認識処理では、あらかじめデータベースの形で音声認識装置内に保存された単語群の中から、話者が発声した単語に該当するものを識別し、結果を話者に返すことが基本的な動作となる。以下、図面を参照しながら従来の特定話者方式の音声認識装置についておおまかな動作説明を行う。図９は従来の特定話者方式の音声認識装置の構成図、図１０は図９中の音声認識処理部の詳細図、図１１は図１０中の単語音響データ格納部の詳細図である。話者の発声した単語は、マイク１で電気信号へ変換され、信号処理部２にて後の処理に適した形式の音声信号へ変換されて音声認識処理部４へ送られる。音声認識処理部４内の音響処理部６はこの音声信号から音響的な特徴量を抽出し、単語識別部８では入力された音響データにもっとも一致するものを単語音響データ格納部７に保持されている音響データの中から探し出す。この結果一致した音響データに関連づけられた単語識別子が識別情報として信号処理部２へ戻され、それによって信号処理部２は話者の発声した単語を認識でき、適切な処理制御を実施する。以上が話者不特定および話者特定の音声認識方式に共通する基本的な認識処理の流れであるが、両方式の基本的な相違点

は単語音響データ格納部７の単語音響データの生成方法にある。前述したように話者特定方式においては単語音響データはトレーニングによって生成される。したがって、装置の初期状態では単語音響データは未定義の状態であるため、音声認識処理の前にこのトレーニングが必須となる。トレーニングとは、話者が認識対象であるすべての単語について発声を行い、それを単語音響データ格納部に登録する処理である。トレーニングにおいて、話者は発声した特定の認識対象の単語はマイク１により入力され信号処理部２によって音声信号に変換されるが、このとき個々の認識対象単語を区別するための単語識別子が付加される。音声認識処理部４ではこの音声信号を音響処理部６で音響データに変換し、単語識別子とともに単語音響データ格納部７へ供給する。単語音響データ格納部７では、この音響データと単語識別子が互いに関連付けて格納される。こうして全ての音声認識対象の単語に対して同様のトレーニングを繰り返すことにより初めて音声認識が可能になる（図７参照）。一方、話者不特定方式においては、単語音響データ格納部内の単語音響データの作成には話者の発声を必要としないため、話者の音声認識動作の前にあらかじめ設定しておくことが可能であり話者の負担はない。ただし、話者が新しく認識単語を追加するためには単語音響データの生成に複雑な計算を要することから小規模な音声認識装置では実現が難しいという欠点もある。この点で、話者特定方式では音響データの生成は、音声認識処理と共通の音響処理部８によって容易に実現されるので、話者による新規の認識単語の追加が小規模な認識装置でも簡単に実現できるという利点がある。図中の音声合成処理部３はテキストを音声に変換する処理部である。音声認識処理には直接の関連はないが、音声認識機能を備えた装置においては一般的に併用されており、本特許においては、この音声合成機能を積極的に利用することを特徴とするため、説明のために併記している。音として出力したいテキストはテキスト情報として信号処理部２から音声合成処置部３へ送られ、結果としの合成音声信号が返される。この合成音声信号は話者に音声として伝えるためスピーカー１２で出力される。

【発明が解決しようとする課題】前述のように特定話者方式の音声認識装置ではトレーニング作業が必要であり、認識対象の単語が多いシステムにおいては話者のトレーニングに要する時間も大きく、その話者への負担がシステムの利便性を低下させてしまっていた。一方、話者不特定方式の装置では、組み込み機器に代表される小規模な装置において新規の認識単語の追加が困難であるためシステムの拡張性や応用性が制限されてしまうし、認識単語の追加が可能になるほど装置の計算能力を高めてしまうと、コストが増大して小規模システムには適用できない等の課題があった。

【課題を解決するための手段】本発明は上記従来の課題を解決するために、低コストで実現できる特定話者の音声認識装置を基本として、一般的に音声認識と併用されることの多い音声合成機能を利用し、それから生成される音声信号を話者による発声の代わりにトレーニングに使用することを特徴とする音声認識装置である。話者に負担となるトレーニングを装置内部で自動的に行うことで、トレーニングのない不特定話者方式の音声認識装置と同様の利便性を提供することができる。

【発明の実施の形態】本発明の請求項１に記載の発明は、音声入力装置であるマイクと、音声出力装置であるスピーカーと、キーボードなどの入力装置と、認識結果を表示する表示装置と、前記マイク、前記スピーカー、前記入力装置および前記出力装置が接続され、音声認識装置全体の処理制御を実施する信号処理部と、前記信号処理部よりテキスト情報を入力され合成音声信号を出力する音声合成処理部と、前記信号処理部より入力された音声信号を、内部で保持している複数の音響データと比較し、その一致結果を音声認識の結果として前記信号処理へ出力する音声認識処理部を備えた音声認識装置において、前記音声合成処理部からの合成音声信号を音響データである合成音音響データに変換する合成音声変換部を備えることで、合成音音響データを従来は話者によるトレーニングで生成されていた音響データの代わりに音声認識処理部内に保持し音声認識に用いることにより、話者の負担になるトレーニングを必要としないことを特徴とし、トレーニングのない不特定話者方式の音声認識装置と同様の利便性を提供することができる。本発明の請求項２に記載の発明は、請求項１の音声認識処理部において、合成音声変換部より出力された合成音音響データかあるいは音響処理部より出力された音響データを選択する音響データ選択部を有することにより、単語音響データ格納部に対して装置の初期時は合成音声音響データを格納しておき、ある認識単語を認識した場合には該当単語の発話音声に相当する音響データで合成音音響データを置き換えることで初回以降は実発音での音響データに対する認識を可能とすることを特徴とし、合成音声と実発音が異なる場合にも音声認識率の低下を防ぐことが出来るという作用を有する。本発明の請求項３に記載の発明は、請求項２の単語音響データ格納部において、前記合成音声変換部より出力された合成音音響データと、前記音響処理部より出力された話者の発話音声の音響データを両方を保持し、両者のうちのいずれかに一致したときに該当する単語の識別情報を出力することを特徴とするものであり、音声認識率のさらなる向上を図ることができるという作用を有する。本発明の請求項４に記載の発明は、請求項１の単語音響データ格納部において、一つの単語に対して複数の発音の仕方を前記合成音声変換部より入力し格納する構成を持つことを特徴とするものであり、話者がいずれの発音を行った場合でも、該当する単語を正しく認識できるという作用を有する。

本発明の請求項５に記載の発明は、請求項４の単語音響データ格納部において、話者が該当する単語を発声し、一致した音響データを残して他の音響データを削除することを特徴とするものであり、次回から不要な識別処理を省略し、認識処理をより高速に行うことが出来るという作用を有する。本発明の請求項６に記載の発明は、請求項５の単語音響データ格納部において、ある単語に対する複数の合成音音響データのそれぞれに話者の発生が一致した頻度を保持する機構を追加する。該当単語の認識が行われた際に、スレッシュレベル以下の一致頻度の音響データのみを削除することを特徴とするものであり、次回から不要な識別処理を省略し、認識処理を高速に行うことが出来るという作用を有する。本発明の請求項７に記載の発明は、請求項１の発明において、個々の識別対象単語に対する合成音響データを音声認識処理部に保持する際に、話者に対して合成音を前記スピーカーより再生し、話者の意図する合成音であるかどうかを確認し、意図しない場合に限り話者によるトレーニング手続きを行うことを特徴とするものであり、話者がトレーニングを実施する頻度を少なくすることができるという作用を有する。以下、本発明の実施の形態について、図面を参照しながら説明する。

（実施の形態１）図１に本発明を適用した音声認識装置の基本的な構成例を示す。図２は図１における音声認識処理部１６の内部構成図である。図１においては、基本的構成は図９の従来の一般的な特定話者方式音声認識装置と同一であるが、合成音声変換部１９を備えることを特徴としている。図１において、１３は音声入力装置であるマイクであり、話者が発声した音声を電気信号へ変換する。１８はスピーカー、２３は表示部である。１７は話者が認識結果の確認を行うためのキー入力や装置全体を制御するための入力装置である。１４は信号処理部、１５は音声合成処理部、１６は音声認識処理部、１９は合成音声変換部である。図２に示す音声認識処理部１６内において、２０は入力音声信号から音響的な特徴量を抽出する音響処理部である、２１は電話帳にあるすべての相手先の名前を認識単語としてそれぞれの単語音響データを保持する単語音響データ格納部である。２２は入力された音響データにもっとも一致するものを単語音響データ格納部２１の中から探し出す単語識別部である。話者の発声した単語はマイク１３で電気信号へ変換され、信号処理部１４へ入力される。信号処理部１４では入力された音声信号を音声認識処理部１６での処理に適した形式の音声信号へ変換する。図２に示す音声認識処理部１６内において音響処理部２０は、信号処理部１４が出力する音声信号から音響的な特徴量を抽出し音響データとして単語識別部２２へと出力する。単語識別部２２では入力された音響データにもっとも一致するものを単語音響データ格納部２１に保持されている音響データの中から探し出す。この結果一致した音響データに関連づけられた単語識別子が識別情報として信号処理部１４へと戻される。図１において、信号処理部１４では音声認識の結果である識別情報によって話者の発声した単語を認識し、それに基づいて装置の適切な処理制御を実施したり、表示装置２３を介して話者に認識結果をフィードバックする。　ここで、本発明の音声認識装置を電話機に適用し、電話機における音声による電話帳検索処理のために、トレーニングに代わって実施される音声合成による単語音響データ格納部への音響データの登録処理について説明する。電話機おいて電話帳は話者が通信を行う相手の名前とその相手の電話番号やメールアドレス等の個人情報を電話内部で保持している一種のデータベースである。話者はこの電話帳に相手先の情報を登録しておけば、毎回電話番号を入力することなく容易に電話をかけることができる。音声認識装置を組み込んだ電話機においては、話者が相手の名前を発声することで自動的に相手の電話番号を電話帳から検索し電話をかけるという、いわゆるボイスダイアリング機能として利用されることが多い。ボイスダイアリング機能の実現のためには、電話帳にあるすべての相手先の名前を認識単語としてそれぞれの単語音響データを単語音響データ格納部２１に保持しておかねばならない。このとき単語識別子として、テキスト形式の相手先の名前か、あるいは電話帳におけるエントリ番号が保持される。従来の特定話者の音声認識装置では音響データは話者が発声する必要があるため、電話帳にあるすべての相手先の名前を発声する必要があった。本発明ではそれに該当する処理を合成音声信号を生成する音声合成処理部１５と、合成音声を音響データに変換する合成音声変換部１９を用いて話者に暗黙的に実行する。具体例として、電話帳に１００件の名前を以下の様に新規に登録する場合を考える。この電話帳のデータは信号処理部１４の内部メモリに保持されるものであるが、先に述べたようにボイスダイアリングのためには、単語音響データ格納部２１にも音響データを登録する必要がある。従来は話者が相手先名前「adam」およびその電話番号「１１１－２２２２」を入力装置２３のキーより入力した場合、これらは信号処理部１４内の電話帳のエントリ１に登録されるが、単語音響データ格納部２１の入力として必要な音響データを音響処理部２０で生成するためにマイク１３によって「adam」を実際に発声する。同時にこれに対応するエントリ番号「１」が単語識別子として音響データに関連付けられ単語音響データ格納部２１に登録される。この発声を伴う音響データの登録には、話者の発声自体に要する時間に加えて、登録処理との発声タイミングをとる時間も必要であり、１件あたり数秒から数十秒の時間を要し、その手順を１００回繰り返す必要があったので話者への負荷は大きかった。本発明においては、従来と同様に相手先名前と電話番号をキー入力した直後に、信号処理部１４の制御によって相手先名前「adam」がテキスト情報とし

て音声合成処理部１５に送られる。合成音声処理部１５では「adam」に相当した標準的な発声音である合成音声信号「アダム」が生成される。この合成音声信号は合成音声変換部１９によってその特徴量データである合成音音響データが生成される。この合成音音響データは信号処理部１４からの単語識別子「１」とともに音声認識処理部１６内にある単語音響データ格納部２１に保存される。こうして話者による発声が伴わないために、音響データ登録の一連の処理は数秒内で自動的に実施される。使用者が相手先名前と電話番号のキー入力を１００回繰り返すことによって単語音響データ格納部２１への音響データの登録処理は完了する。この場合、合成音音響データにはマイクからの音声を変換して作成した音響データと異なりマイク周囲の雑音の混入による影響もない。使用者は相手先名前と電話番号のキー入力を登録件数分行うが、トレーニングのためにマイクに向かって発声する必要がなく、登録処理との発声タイミングをとる時間も必要ないので、使用者への負担は少ない。表１は本願発明の電話機の電話帳の構成例を示す。

【表１】

エントリ番号	相手先名前	電話番号
1	Adam	１１１－２２２２
2	Henry	１１０－２３３３
3	John	１２３－３３４４
・・・	・・・	・・・
１００	Tom	２２２－４４５５

電話機の電話帳の構成例

登録完了後、使用者が例えば「john」に電話をかけたい場合、相手の名前である「ジョン」をマイク１３から入力する。この音声は信号処理部１４を経由して音声信号として音声認識処理部１６に送られる。音響処理部２０はこの「ジョン」を音響データに変換し単語識別部２２へ送る。単語識別部２２はこの音響データを単語音響データ格納部の１００件の音響データと比較し、結果として一致したエントリの単語識別子「３」を識別情報として信号処理部１４へ返す。信号処理部１４では内部の電話帳データを検索し、識別情報である「３」からエントリ番号３の「john」の電話番号「１２３－３３４４」を得ることが出来る。通常はこの電話番号に対して話者の確認を行った後、電話がかけられる。以上のように本実施の形態によれば、低コストの特定話者方式の音声認識装置でありながら、話者に負担となるトレーニングを装置内部で自動的に行うことで、見かけ上はトレーニングのない不特定話者方式の音声認識装置と同様の利便性の音声認識装置を得ることができるという効果が生じる。

（実施の形態２）実施の形態１で説明したように単語音響データ格納部の音響データをすべて合成音音響データとすることによって、話者をトレーニングから解放することができた。しかし、この方法では、方言等の影響によって話者の発声する単語が音声合成変換部で生成される標準的な発音と非常に異っている場合には、話者の単語が認識が困難になることが考えられる。実施の形態２はこの課題を解決するために、基本的な構成は形態１と同様であるが、図３に示すように音声認識処理部に音響データ選択部２７を設けることによって、単語音響データ格納部２５に格納する認識単語の音響データを合成音音響データかあるいは音響処理部２４からの音響データかの選択を可能とした。以下に形態１と同様の電話機でのボイスダイアリングを例にしてその動作を説明する。この発明においては１００件の電話帳登録に伴う単語音響データ格納部２５への音響データ登録の際には、音響データ選択部２７では合成音音響データが選択される。したがって、登録完了時点では単語音響データ格納部２５の登録内容は形態１と同一となる。しかし、エントリ番号２の「Henry」に対する合成音声信号が「ヘンリー」であるのに、実際の読みが「アンリ」だった場合を考える。ボイスダイアリングにおいて話者が発する「アンリ」に対して単語識別部２６が正しくエントリ番号「２」を識別情報として信号処理部に返す割合（いわゆる音声認識率）は、単語音響データ格納部２５に正しい読みである「アンリ」が登録されている場合よりも小さくなってしまう。本発明では、このような合成音声と実際の発声音との相違がある場合の音声認識率の低下を防ぐために、単語識別部２６からの識別情報に基づいて、単語音響データ格納部２５の該当音響データを音響処理部２４が出力する音響データに置き換える機構を設ける。この場合には、話者が「アンリ」と発声し、単語識別部２６から識別情報としてエントリ番号「２」が信号処理部に戻された時点で、音響処理部２４に保持されていた「アンリ」に相当する音響データが音響データ選択２７によって選択され、同時に信号処理部からは単語識別子「２」が単語音響データ格納部２５に入力される。これらのデータから単語音響データ格納部のエントリ２は「ヘンリー」から「アンリ」に相当する音響データに置き換えられる。この処理は話者に対しては暗黙的に実行されるため、話者の負担は生じない。こうして、実際の音声がその合成音声と多少異なる場合でも、識別された単語に対する音響データは実際の話者の発声音に相当する音響データに常に更新されていくため、合成音声と実発音の違いに基づく永続的な音声認識率の低下を防ぐ効果がある。

（実施の形態３）形態２においては認識が行われた場合には、該当単語の音響データを発声された音響データに置き換えて、合成音声が実発音と異なる場合の音声認識率の低下を防ぐことを目的とした。しかし、置き換えとして登録される音声が常に適切なものとは限らない。たとえば、発音の異なる複数の話者が音声認識装置を共用

している場合や、同一話者の発声においても周囲のノイズが混入した場合など、一度特殊な発声の音響データが該当単語の音響データとして単語音響データ格納部に登録されてしまうと、標準的な発声を行っても音声認識率が低下してしまうことが考えられる。実施の形態３ではこの課題を解決するために、図４に示すように単語音響データ格納部を認識対象単語毎に合成音音響データ２９と発声音響データ３０を保持できるようにする。これにより標準的な合成音声に近い発声においても、また話者独特の標準的でない発声の場合においても、いづれかの音響データに一致することで音声認識率の低下を防ぐことができる。電話帳の先例を用いて動作を説明する。エントリ番号２の「Henry」に対する合成音声信号が「ヘンリー」であるのに、実際の読みが「アンリ」だった場合を考える。まず、単語音響データ格納部において、電話帳の登録時に単語識別子２８にエントリ番号「２」が格納され、それに関連する合成音音響データ２９に「ヘンリー」が格納される。この後、実際の話者の発生「アンリ」によって認識が成功し、エントリ「２」が識別情報として信号制御部に返された時に「アンリ」がエントリ「２」に関連する発生音響データ３０へ格納される。この一連の処理は形態２において音響データの書き換えが起こらない点を除いて同一である。こうして、電話帳の単語「Henry」について２つの音響データが単語音響データ格納部に保持されることになる。こうして次のボイスダイアリングにおいて話者が「アンリ」と発声した場合においても、別の話者が「ヘンリー」と発声した場合においても、どちらも音声認識率を低下させることなく音声認識処理が実施できるという効果がある。発声音響データ３０については、形態２と同様に、エントリ「２」への認識が成功するたびに話者の発声に対応する音響データへと更新される。

（実施の形態４）実施の形態１においては1つの認識単語に対して登録できる音響データは1つであった。しかし、電話帳登録の例での「Henry」のように、あらかじめ複数の発音が存在することがわかっている場合には、電話帳の登録時にその全てを登録するほうが音声認識率を上げることができる。実施の形態４は、基本的な構成は形態１と同一であるが、図５にあるように1つの認識単語に対して複数の合成音音響データを登録できる単語音響データ格納部を備えることを特徴とし、話者の発声がいづれかの合成音音響データに一致した場合には、該当単語の認識を可能にしたものである。たとえば、「Henry」の例では、単語識別子３１にエントリ番号「２」が格納され、それに関連する合成音音響データＡ３２に「ヘンリー」に該当する音響データが、合成音音響データＢ３３に「アンリ」に該当する合成音音響データが格納される。この２つの合成音音響データは合成音声変換部において一般的な読み方の知識を用いて自動的に生成される。こうして、話者が「ヘンリー」あるいは「アンリ」と発声した場合においても認識率を低下させることなく正しく「Henry」が認識される。

（実施の形態５）実施の形態４において、1つの認識単語に対して複数の合成音音響データを持つことを可能にすることで異なる発声に対する認識率の低下を改善できた。しかし、認識の際に単語識別部が比較する音響データが増大するため音声認識処理の時間を増加させるという問題がある。実施の形態５では、単語音響データ格納部内のそれぞれの単語の合成音音響データについて、認識単語のうち話者の発声にもっとも一致した合成音音響データを残して、他の合成音音響データを削除する機構を備えることによって、次回の認識処理から不要な識別処理を省略し、認識処理をより高速に行うことを特徴とする。

（実施の形態６）実施の形態５では、認識された単語の複数の合成音音響データのうち、もっとも一致したものを除いて他の全てが削除された。しかし、ある認識単語に対して１つ以上の発音が同程度に発生することも考えられ、この場合には残された発音以外の音声認識率が低くなってしまう。実施の形態６ではこれを解決する手段として、図６の単語音響データ格納部において、各単語の複数の合成音音響データそれぞれに対して該当単語の認識が行われた際に一致した頻度情報を記録する機構を追加する。この頻度情報が一定のスレッシュレベルを下回った場合に限り、その合成音音響データを削除する。このようにして、発声頻度の少ない合成音音響データを音声認識のたびに徐々に削除することで、発声可能性のある合成音音響データを残すことで音声認識率の低下を防ぎつつ、不要な音響データによる音声認識時間の増大を防ぐ効果がある。電話帳における「Henry」の例を用いて動作を説明する。図６の単語音響データ格納部において、電話帳登録時に単語識別子３５にエントリ番号「２」が格納され、それに関連する合成音音響データＡ３６に「ヘンリー」に該当する音響データ、合成音音響データＢ３８に「アンリ」に該当する合成音音響データ、合成音音響データＣ４０に「ヘンリ」に該当する合成音音響データの３種類の可能性のある発音が格納されたとする。同時にこの登録時にはそれぞれの合成音音響データの頻度情報として、初期値１０が頻度情報Ａ３７、頻度情報Ｂ３９および頻度情報Ｃ４１に格納される。まず音声認識において話者が「ヘンリー」と発声し「Henry」が認識された場合には、「ヘンリー」に該当する頻度情報Ａの数値は１加算されるが最大値は初期値１０を超えないため１０のままとなり、「ヘンリー」以外の頻度情報すなわち頻度情報Ｂおよび頻度情報Ｃの数値が１減算され、それぞれ９となる。この「Henry」に対する認識が「ヘンリー」の発音でさらに７回続いた場合には、頻度情報Ａ、Ｂ、Ｃはそれぞれ１０、２、２となる。次に「アンリ」による認識が起こった場合、頻度情報Ａ、Ｂ、Ｃはそれぞれ９、３、１となる。次に「ヘ

ンリー」による認識が起こった場合、頻度情報Ａ、Ｂ、Ｃはそれぞれ１０、２、０となる。この時点で「ヘンリ」に対応する頻度情報はスレッシュレベル１を下回り音声音響データＣ４０は削除される。これ以降は「ヘンリ」に対する比較処理は行われない。このようにして発声頻度の小さい合成音音響データのみが音声認識が進むにつれて削除されることになる。

（実施の形態７）実施の形態１においては、認識単語の音響データとして合成音音響データを登録するが、話者の実際の単語の読みと合成音声がまったく異なる場合には、該当単語の認識が困難になるという課題がある。この問題を解決するために、発明の構成は形態２と同様であるが、登録する合成音声信号を音声合成処理部が生成した際に、その信号を合成音声変換部に入力すると同時に信号処理部へも返し、信号処理部ではスピーカーを用いて話者にフィードバックした後、話者からの確認入力を入力装置を通じて得るような機構を設ける。これによって、話者は合成音声が意図した読みに近い場合にはＯＫの確認入力を行い、合成音声が意図する読みと全く異なる場合にはＮＧの確認入力を行うようにする。確認入力がＮＧの場合には、信号処理部においてマイクを通じて該当単語に対する話者の発声を入力するようにし、その音声信号を音響処理部、音響データ選択部を単語音響データ登録部に格納するような制御を行う。以上のようにして、認識単語の音響データ登録時に話者への確認処理を行うことによって、話者のトレーニングによる処理を最小限にし、かつ意図しない合成音声が登録されて音声認識率が低下することを防ぐことができる。

【発明の効果】本発明は低コストで実現できる特定話者の音声認識装置を基本として、一般的に音声認識と併用されることの多い音声合成機能を利用し、それから生成される音声信号を話者による発声の代わりに装置内部で自動的に行うことで、見かけ上は話者に負担となるトレーニングのない音声認識装置を提供することができる。

【図面の簡単な説明】

【図１】本発明の音声認識装置を搭載した電話機の基本的構成を示すブロック図

＊【図２】本発明の実施の形態１における音声認識処理部の構成を示すブロック図

【図３】本発明の実施の形態２における音声認識処理部の構成を示すブロック図

【図４】本発明の音声認識装置に必要な単語音響データ格納部の一例を示すブロック図

【図５】本発明の音声認識装置に必要な単語音響データ格納部の他の例を示すブロック図

【図６】本発明の音声認識装置に必要な単語音響データ格納部の他の例を示すブロック図

【図７】トレーニングを伴う従来の単語音響データの登録処理手順を示すフローチャート

【図８】本特許によるトレーニングを伴わない単語音響データの登録処理手順を示すフローチャート

【図９】一般的な特定話者法式の音声認識装置の構成を示すブロック図

【図１０】同音声認識装置の音声認識処理部の構成を示すブロック図

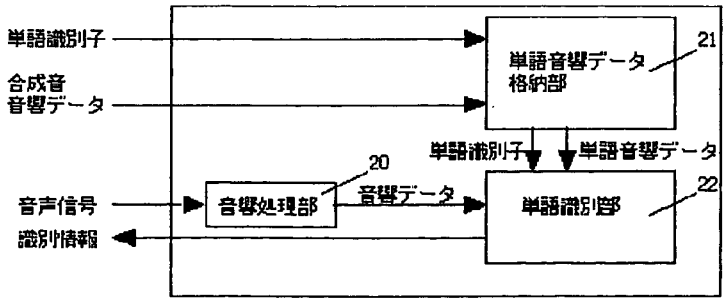
【図１１】同音声認識装置の単語音響データ格納部の構成を示すブロック図

【符号の説明】

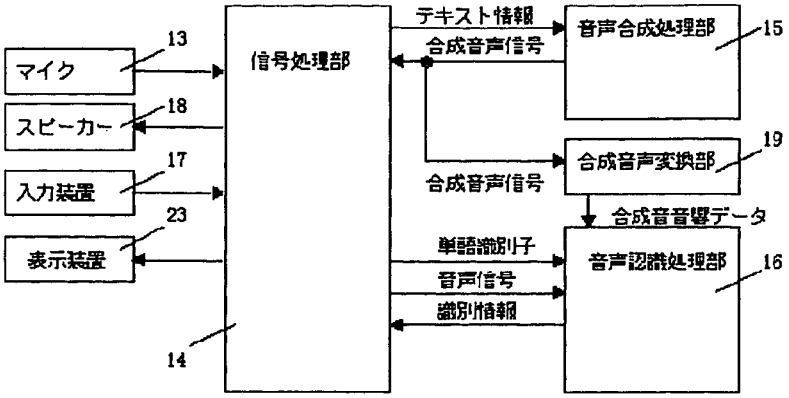
１３　マイク
１４　信号処理部
１５　音声合成処理部
１６　音声認識処理部
１７　入力装置
１８　スピーカー
１９　合成音声変換部
２０　音響処理部
２１　単語音響データ格納部
２２　単語識別部
２３　表示装置
２７　音響データ選択部
２８　単語識別子
２９　合成音音響データ
３０　発声音響データ

＊

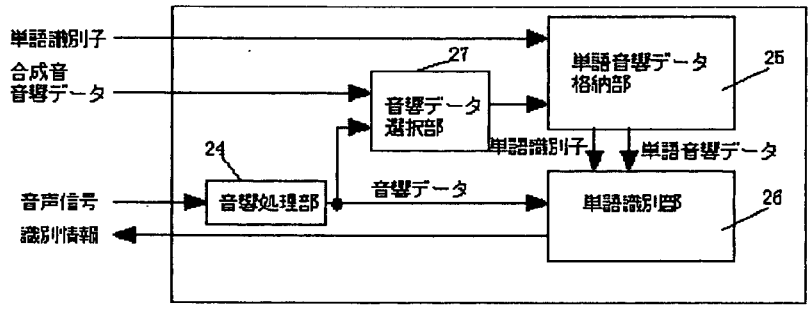
【図２】



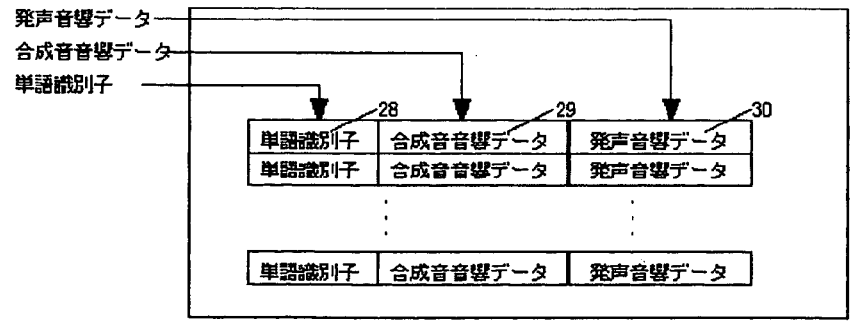
【図１】


マイク
スピーカー
入力装置
表示装置
信号処理部
13
18
17
23
14
テキスト情報
合成音声信号
音声合成処理部
15
合成音声信号
合成音声変換部
19
合成音音響データ
単語識別子
音声信号
識別情報
音声認識処理部
16

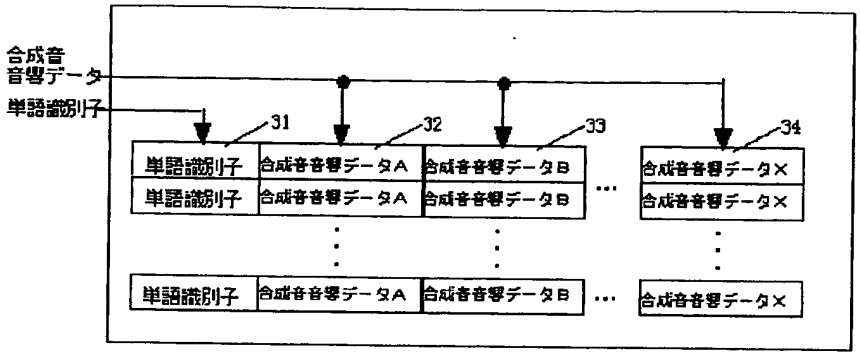
【図３】


単語識別子
合成音
音響データ
音声信号
識別情報
音響処理部
24
27
音響データ
選択部
音響データ
単語識別子
単語音響データ
格納部
25
単語識別子
単語音響データ
単語識別部
26

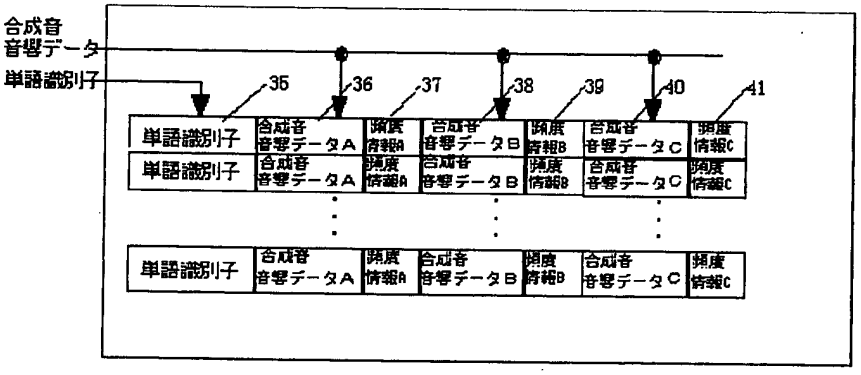
【図４】


発声音響データ
合成音音響データ
単語識別子
28
29
30
単語識別子
合成音音響データ
発声音響データ
単語識別子
合成音音響データ
発声音響データ
単語識別子
合成音音響データ
発声音響データ

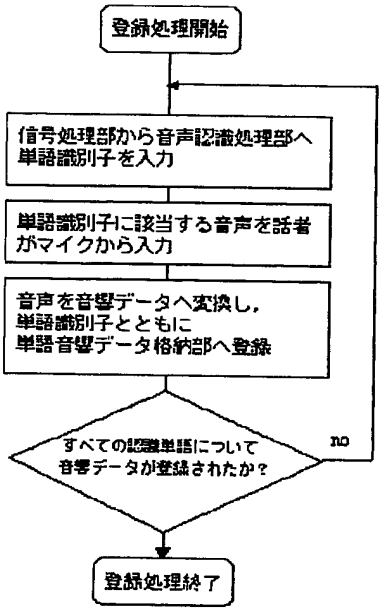
【図５】


合成音
音響データ
単語識別子
31
32
33
34
単語識別子
合成音音響データA
合成音音響データB
合成音音響データX

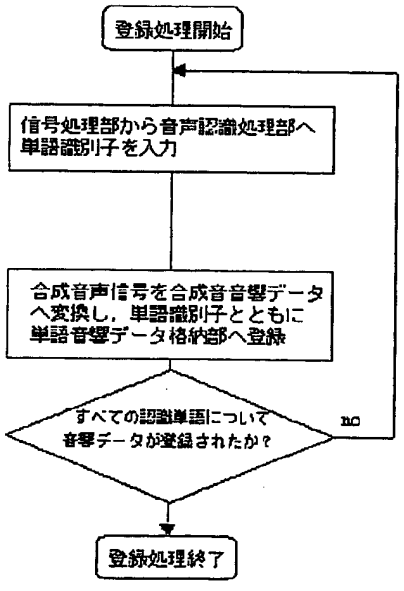
【図６】


合成音
音響データ
単語識別子
35
36
37
38
39
40
41
単語識別子
合成音
音響データA
頻度
情報A
合成音
音響データB
頻度
情報B
合成音
音響データC
頻度
情報C

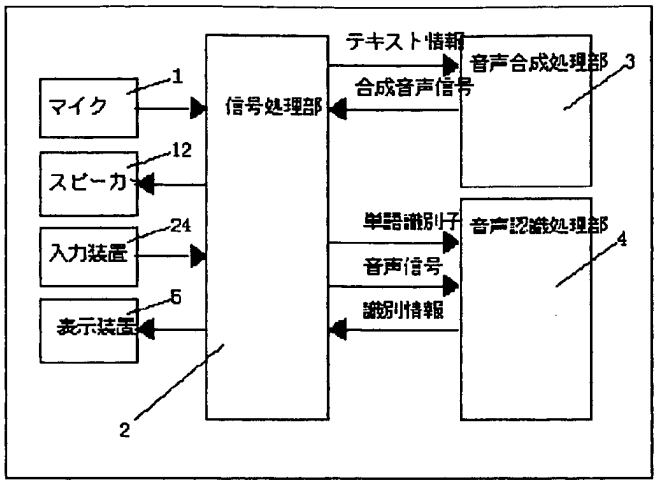
【図７】


登録処理開始
信号処理部から音声認識処理部へ
単語識別子を入力
単語識別子に該当する音声を話者
がマイクから入力
音声を音響データへ変換し，
単語識別子とともに
単語音響データ格納部へ登録
すべての認識単語について
音響データが登録されたか？
no
登録処理終了

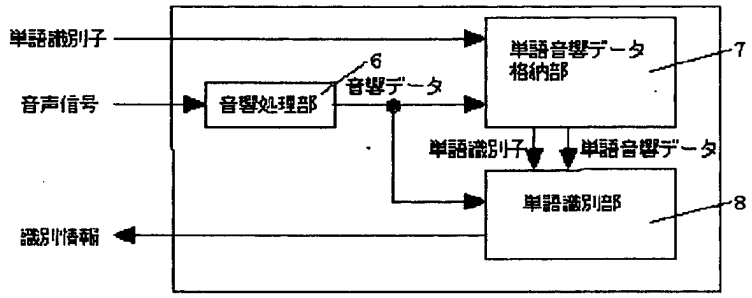
【図８】


登録処理開始
信号処理部から音声認識処理部へ
単語識別子を入力
合成音声信号を合成音音響データ
へ変換し，単語識別子とともに
単語音響データ格納部へ登録
すべての認識単語について
音響データが登録されたか？
no
登録処理終了

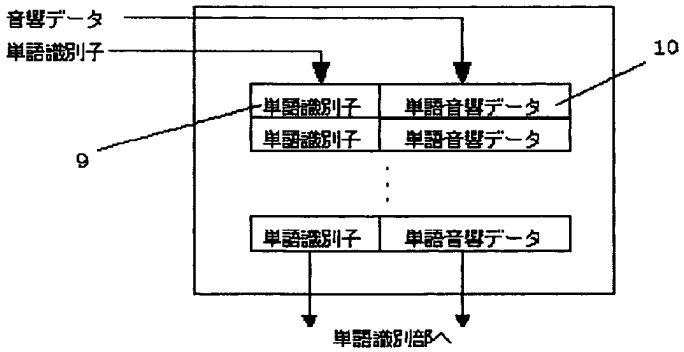
【図９】


マイク
1
信号処理部
テキスト情報
音声合成処理部
3
合成音声信号
スピーカ
12
入力装置
24
単語識別子
音声認識処理部
4
音声信号
表示装置
5
識別情報
2

【図１０】


単語識別子
単語音響データ
格納部
7
音声信号
音響処理部
6
音響データ
単語識別子
単語音響データ
単語識別部
8
識別情報

【図１１】


音響データ
単語識別子
10
単語識別子
単語音響データ
9
単語識別子
単語音響データ
単語識別子
単語音響データ
単語識別部へ

フロントページの続き

(72)発明者　立山　雅一
大阪府門真市大字門真1006番地　松下電器産業株式会社内
(72)発明者　西田　博人
大阪府門真市大字門真1006番地　松下電器産業株式会社内
(72)発明者　黒木　義明
大阪府門真市大字門真1006番地　松下電器産業株式会社内
(72)発明者　西岡　靖幸
大阪府門真市大字門真1006番地　松下電器産業株式会社内
(72)発明者　五島　龍宏
大阪府門真市大字門真1006番地　松下電器産業株式会社内

Ｆターム(参考)　5D015 AA03 HH04 KK04
5D045 AB30